Report of Findings

Exploratory Data Analysis – Titanic Dataset

Dataset Overview

- **Total Records**: 887 passengers
- **Features**: 8 original columns
- **No missing values** after preprocessing
- **New Features Created**:
  - FamilySize: Combined family aboard
  - IsAlone: Binary feature (1 = alone)
  - AgeBin, FareBin: Categorical bins for grouped analysis

Statistical Highlights

Feature	Mean	Min	Max	Insight
Age	29.47	0.42	80.0	Majority of passengers are young adults
Fare	₹32.30	0.00	512.33	Fare distribution is skewed with a few very high values
Survived	38.6%	0	1	Only about 39% of passengers survived
Pclass	Mean: 2.3	Range: 1–3	Most passengers are from class 3	

Key Findings from Visuals (Not shown here)

1. **Survival Rate by Gender**
   - Females had a significantly higher survival rate than males.
2. **Survival by Class**
   - First-class passengers had the highest survival rate.
3. **Fare & Age Distribution**
   - Fare is positively correlated with survival (rich survived more).
   - Children (age < 10) had better chances of survival.

4. **Family Size Effect**
   - Passengers with large families or no family (IsAlone = 1) had lower survival rates.

5. **Heatmap (Correlation)**
   - Fare, Pclass, and Sex showed the strongest relationships with Survived.

Conclusion

The EDA reveals significant patterns between survival and socio-economic factors (class, fare), gender, and family size. Feature engineering added interpretability for later modeling stages. This analysis builds a strong foundation for predictive modeling.